TABLE 1

NOMBRES TIRÉS AU HASARD

1) Extrait d'une page de la table de Tippett

7366	3899	3862	0902	8484	0860	8388	6686
2968	9888	4487	4562	5842	9808	5051	2674
4528	9223	8277	2057	5139	8591	8750	3416
9208	5446	6277	6716	1989	8260	6303	4672
5215	0374	2196	6662	8772	9055	4198	2200
6219	6942	1918	0880	8673	4847	0721	3341
0014	5871	0562	3986	3122	5596	5171	1130
6179	9227	2302	0811	9167	0840	7558	4442
3380	2006	3250	9732	3506	8100	8827	1243
4139	9866	5974	9813	0988	1048	2065	7686
....	....	....	....	....	....	....	....

Mode d'emploi: A partir d'un chiffre pris au hasard, on suit une ligne ou une colonne, ou une diagonale quelconque.

CHA

DISTR D'ÉCHAN

EXE

On considère une population de 1 cm) des éléments de cette populati

On sait que, pour la population, la t On tire au hasard des échantillons

1°) Quelle loi suit la taille moyenn l'écart-type de cette distribution d

2°) Déterminer la probabilité pour la population, la taille moyenne s

3°) Déterminer la probabilité pou la population, la taille moyenne s

4°) Déterminer, enfin, la probabil issu de la population, la taille mo

1°) *Loi de la taille moyenne de l*

Soient $\bar{x}_i$ et $\sigma_{\bar{x}}$ les paramètres mo l'échantillon (n = 100) étant gra

$$\bar{x} = m = 170$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{10} = 2$$

ainsi : $\bar{x}_i \rightarrow N(170 ; 2)$

2°) $P(\bar{x} < 165) = \pi\left(\frac{165 - 170}{2}\right.$

3°) $P(\bar{x} > 172) = 1 - \pi\left(\frac{172 -}{}\right.$

4°) $P(168 < \bar{x} \leq 175) = \pi\left(\frac{17}{}\right.$

$= \pi(2.5$

EXERCICE N° 109

La consommation de sucre par ménage et par an d'une population donnée est voisine de 50 kg avec écart-type de 5 kg.

On prélève au hasard un échantillon de 100 personnes.

1°) A quelle loi obéit la quantité $\bar{x}$ consommation moyenne de sucre de l'échantillon ? Indiquer les paramètres.

2°) Un échantillon donne $\bar{x} = 52$ kg. Peut-on considérer cet échantillon comme représentatif de la population ?

(Licence 2ᵉ année. Sc. Eco. Octobre 1975)

Solution

1°) La taille de l'échantillon étant grande, $\bar{x}$ suit une loi normale de paramètres :

$E(\bar{x}) = m = 50$ kg

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{10} = 0{,}5$$

m et σ étant respectivement la moyenne et l'écart-type de la population.

2°) Un échantillon qui donne $\bar{x} = 52$ kg est-il représentatif ? On cherche si $\bar{x} = 52$ est contenu dans un intervalle à 95 % et à 99 %.

Pour un intervalle à 95 %, la borne supérieure est :

$50 + 2 \times 0{,}5 = 51$ kg

Pour un intervalle à 99 %, la borne supérieure est :

$50 + 3 \times 0{,}5 = 51{,}5$ kg

Ainsi il y a 99 % de chances que $\bar{x} = 52$ kg se trouve à l'extérieur de l'intervalle $m \pm 3\sigma$. Par conséquent un échantillon qui fournit une telle moyenne n'est pas représentatif car 52 est loin de 50.

EXERCICE N° 110

On considère la quantité $\chi^2 = n \cdot \dfrac{\sigma_e^2}{\sigma^2}$ où σ_e^2 représente la variance de l'échantillon (quantité aléatoire) et σ^2 la variance de la population (quantité certaine).

En sachant que $V(\chi^2) = 2(n-1)$, donner une expression de la variance de σ_e^2 puis une expression de la variance de s^2.

(Licence 2ᵉ année. Sc. Eco. Juin 1975)

Solution

On sait que :

$$\chi^2 = n \cdot \frac{\sigma_e^2}{\sigma^2}$$

$$V(\chi^2) = V\left(n \cdot \frac{\sigma_e^2}{\sigma^2}\right)$$

(n et σ^2 étant constants peuvent sortir de l'opérateur-variance en s'élevant au carré)

$$V(\chi^2) = \frac{n^2}{\sigma^4} V(\sigma_e^2) \Rightarrow V(\sigma_e^2) = \frac{\sigma^4}{n^2} V(\chi^2)$$

d'où $V(\sigma_e^2) = \dfrac{2(n-1)\,\sigma^4}{n^2}$

Remarque : si $n \to \infty$ alors $(n-1) \to n$; $V(\sigma_e^2) = \dfrac{2\sigma^4}{n} \to 0$

par ailleurs

$$V(s^2) = V\left(\sigma_e^2 \cdot \frac{n}{n-1}\right) = \frac{n^2}{(n-1)^2} V(\sigma_e^2)$$

$$= \frac{n^2}{(n-1)^2} \times \frac{2(n-1)\,\sigma^4}{n^2}$$

$$= \frac{2\sigma^4}{n-1}$$

Remarque : si $n \to \infty$; $V(s^2) \to 0$

EXERCICE N° 111

On considère $\chi^2 = n \cdot \dfrac{\sigma_e^2}{\sigma^2}$

Un échantillon de 15 valeurs donne $\sigma_e^2 = 4$

Déterminer « a » tel que $P(\sigma^2 > a) = 0{,}10$

(Licence 2ᵉ année. Sc. Eco. Octobre 1975)

Solution

$$\left.\begin{array}{l} \chi^2 = n \cdot \dfrac{\sigma_e^2}{\sigma^2} \\ P(\sigma^2 > a) = 0{,}10 \end{array}\right\} \Rightarrow P\left[n \cdot \frac{\sigma_e^2}{\chi^2} > a\right] = 0{,}10$$

$$P\left[\chi^2 < n \cdot \frac{\sigma_e^2}{a}\right] = 0{,}10$$

ou bien

$$P\left[\chi^2 > n \cdot \frac{\sigma_0^2}{a}\right] = 0,90$$

$$\left.\begin{array}{l} n = 15 \Rightarrow \nu = 14 \\ 1 - \alpha = 0,10 \Rightarrow \alpha = 0,90 \end{array}\right\} \Rightarrow \chi^2 = 7,79 \text{ (table de khi-deux)}$$

$$n \cdot \frac{\sigma_0^2}{\sigma^2} = 7,79 \rightarrow a = \frac{4 \times 15}{7,79} = 7,702$$

EXERCICE N° 112

Au bout de 3 mois, le directeur commercial d'une entreprise estime raisonnable et légitime d'admettre que la distribution quotidienne des ventes d'un produit X obéit à une loi normale de moyenne m = 500 et d'écart-type σ = 49. Le prix unitaire du produit est de 80 centimes ; le bénéfice obtenu par la vente d'une unité est de 15 centimes.

1°) Appelons $x_1, x_2, \ldots, x_6$ le nombre d'unités vendues chaque jour de la semaine. Quelles sont les lois de probabilité et les intervalles d'encadrement des ventes à 0,95 et 0,99 pour :

— *le nombre d'unités vendues pendant la semaine ;*
— *le nombre moyen d'unités vendues par jour pendant une semaine*

2°) *Quelle est la probabilité pour que le bénéfice soit d'au moins 300 dirhams la prochaine semaine en décidant un approvisionnement de 500 unités par jour ?*

Solution

Soit X « le nombre d'unités vendues par jour » ; on sait que :

$X \rightarrow N(500\ ;\ 49)$

1°) *Lois et intervalles de confiance des ventes :*

a) Pendant une semaine :

Soit S les ventes d'une semaine ; les ventes quotidiennes x_i sont indépendantes. D'où, d'après le théorème central limite, S suit une loi normale :

de moyenne : $E(S) = 500 \times 6 = 3\,000$

de variance : $\sigma_s^2 = 6 \cdot \sigma^2 = 6 \times (49)^2 = 14\,406$

d'écart-type : $\sigma_s = 49\sqrt{6} = 120,02$

$S \rightarrow N(3\,000\ ;\ 120)$

b) Moyennes par jour pendant une semaine

Soit $\bar{x}$ les ventes journalières moyennes pendant une semaine

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6}$$

La population étant gaussienne, $\bar{x}$ suit une loi normale de paramètre

$E(\bar{x}) = 500$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{49}{\sqrt{6}} \simeq 20$$

$\bar{x} \rightarrow N(500\ ;\ 20)$

c) Intervalles de confiance :

— *Pour S*

à 0,95 $P[E(S) - t\sigma_s < S \leq E(S) + t\sigma_s] = 0,95 = 2\pi(t) - 1$

$2\pi(t) - 1 = \alpha = 0,95 \Rightarrow t = 1,96$

$P[3\,000 - 1,96 \times 120 < S \leq 3\,000 + 1,96 \times 120] = 0,95$

$P[2\,764 < S \leq 3\,236] = 0,95$

à 0,99 : $t = 2,58$

$P[2\,691 < S \leq 3\,309] = 0,99$

— *Pour $\bar{x}$*

à 0,95 : $P[E(\bar{x}) - t\sigma_{\bar{x}} < \bar{x} \leq E(\bar{x}) + t\sigma_{\bar{x}}] = 0,95$

$P[500 - 1,96 \times 20 < \bar{x} \leq 500 + 1,96 \times 20] = 0,95$

$P[460 < \bar{x} \leq 540] = 0,95$

à 0,99 : $P[448 < \bar{x} \leq 552] = 0,99$

2°) *Probabilité du bénéfice sur une vente de 500 unités par jour*

Soit : B_i : le bénéfice du jour i de la semaine

B_i = Recette - Dépense

Recette = 0,8 S

Dépense = $0,65 \times 500 \times 6 = 1\,950$

$B_i = 0,8\,S - 1\,950$

S suivant une loi normale, B_i suit également une loi normale.

Ses paramètres sont :

$E(B) = 0,8\,E(S) - 1\,950 = 2\,400 - 1\,950 = 450$

$\sigma_B = 0,8\,\sigma_S = 0,8 \times 120 = 96$

$B_i \rightarrow N(450\ ;\ 96)$

$$P(B_i > 300) = P\left(t > \frac{300 - 450}{96}\right) = P(t > -1,56)$$
$$= 1 - \Pi(-1,56)$$
$$= \Pi(1,56) = 0,94$$

Il y a donc 94 % de chances de faire un bénéfice d'au moins 300 dh si on continue à s'approvisionner régulièrement avec 500 unités par jour

EXERCICE N° 113

On admet que la taille X exprimée en cm d'un individu dans une population donnée est une variable aléatoire normale de paramètre $m = 170$ cm et $\sigma = 5$ cm, m et σ désignent la moyenne et l'écart-type.

1°) Calculer la probabilité conditionnelle pour qu'un individu soit plus grand que 170 cm sachant que sa taille est supérieure à 160 cm.

2°) Quelle est la loi de probabilité de la taille moyenne d'un individu dans un échantillon de 100 personnes choisies dans la population considérée ? Donner une précision de la taille moyenne de l'échantillon dans un intervalle à 95 %.

3°) Quel doit être le nombre d'individus de l'échantillon si l'on veut connaître la taille moyenne à 0,5 unité de la valeur correcte, avec une probabilité de 0,95 ?

(Licence 2° année, Sc. Eco. Juin 1976)

Solution

1°) Soit A l'événement $(X > 170)$
et B l'événement $(X > 160)$
X étant la taille de l'individu

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1 - P(X \leq 170)}{1 - P(X \leq 160)}$$

$$= \frac{1 - \Pi\left[\frac{170 - 170}{5}\right]}{1 - \Pi\left[\frac{160 - 170}{5}\right]} = \frac{1 - \Pi(0)}{1 - \Pi(-2)} = \frac{1 - \Pi(0)}{1 - [1 - \Pi(2)]}$$

$$= \frac{1 - \Pi(0)}{\Pi(2)} = \frac{1 - 0,5}{0,9772} = 0,51 \text{ soit } 51\ \%$$

2°) Loi de $\bar{x}$ dans un échantillon de taille 100 :

$\bar{x} \rightarrow N(E(\bar{x}) ; \sigma_{\bar{x}})$

$$E(\bar{x}) = m = 170 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{10} = 0,5$$

Intervalle à 95 % :

$$P\left\{m - t\frac{\sigma}{\sqrt{n}} < \bar{x} \leq m + t\frac{\sigma}{\sqrt{n}}\right\} = 2P(t) - 1 = 0,95$$

$2\Pi(t) - 1 = 0,95 \Rightarrow t = 1,96 \# 2$

d'où

$$P\{170 - 2 \times 0,5 < \bar{x} \leq 170 + 2 \times 0,5\} = 0,95$$

$$P\{169 < \bar{x} \leq 171\} = 0,95$$

3°) Calcul de n

A la limite on a :

$$t\frac{\sigma}{\sqrt{n}} = \varepsilon$$

ε étant l'erreur maximum admise (voir exercice n° 11[illegible] estimation)

$\alpha = 0,95 \Rightarrow t = 1,96 \# 2$

$\sigma = 5$

$$\frac{t^2\sigma^2}{n} = \varepsilon^2 \Rightarrow n = \frac{t^2\sigma^2}{\varepsilon^2}$$

$$n = \frac{4 \times 25}{(0,5)^2} = \frac{4 \times 25 \times 100}{25} = 400$$

n doit être supérieur à 400

EXERCICE N° 114

Le fabricant d'une machine a garanti à son utilisateur que la longueur moyenne des pièces qu'elle fabrique est de 20 cm avec une variance de 4 cm. Pour vérifier si la machine est bien réglée on prélève régulièrement un échantillon dont on calcule la longueur moyenne qu'on compare à la moyenne théorique m avec une probabilité de 95 %.

1°) L'échantillon prélevé est de 100 pièces. Etablir l'intervalle à 95 % pour l'encadrement de la moyenne.

2°) L'échantillon prélevé est de 25 pièces. Calculer le nouvel intervalle à 95 %. Pour cela on considère que les longueurs des pièces sont distribuées sensiblement suivant une loi normale.

3°) L'échantillon prélevé est de 10 pièces et fournit les longueurs suivantes :

22 cm	22	18	24	18
15,5	18	16	24,5	18

Encadrer la moyenne $\bar{x}$ de cet échantillon dans un intervalle à 95 %.

On donne :

$\sum x_i = 200$

$\sum(x_i - \bar{x})^2 = 76,5$

4°) Comparer et commenter :

a) les différences introduites dans les hypothèses des trois questions.

b) les trois intervalles calculés et en déduire s'il existe un lien entre la taille de l'échantillon et sa représentativité.

(Licence 2° année, Sc. Eco. Octobre 1976)

Solution

1°) $n = 100$

La taille de l'échantillon étant grande, $\bar{x}$ suit une loi normale

$$\bar{x} \to N\left(20 \,;\, \frac{2}{10}\right)$$

$$\alpha = 0{,}95 \to t = 2$$

$$P\left[m - 2\frac{\sigma}{\sqrt{n}} < \bar{x} \leq m + 2\frac{\sigma}{\sqrt{n}}\right] = 0{,}95$$

$$m + t\frac{\sigma}{\sqrt{n}} = 20 + (2 \times 0{,}2) = 20{,}4$$

$$m - t\frac{\sigma}{\sqrt{n}} = 20 - (2 \times 0{,}2) = 19{,}6$$

$$P(19{,}6 < \bar{x} \leq 20{,}4) = 0{,}95$$

2°) $n = 25$

La longueur des pièces étant distribuée normalement, on en déduit que $\bar{x}$ suit une loi normale, d'où :

$$P\left[m - t\frac{\sigma}{\sqrt{n}} < \bar{x} \leq m + t\frac{\sigma}{\sqrt{n}}\right] = 2\Pi(t) - 1 = 0{,}95$$

$$P\left[20 - 2 \times \frac{2}{5} < \bar{x} \leq 20 + 2 \times \frac{2}{5}\right] = 0{,}95$$

$$P(19{,}2 < \bar{x} \leq 20{,}8) = 0{,}95$$

3°) $n = 10$

$$\bar{x} = \frac{200}{10} = 20$$

La variance σ^2 doit être estimée par s^2 car la taille de l'échantillon est petite :

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{76{,}5}{9} = 8{,}5$$

$$s = 2{,}91$$

$\bar{x}$ suit une loi de *STUDENT-FISHER* à 9 degrés de liberté. La table de Student donne pour

$$\left.\begin{array}{l} v = 9 \\ \alpha = 0{,}95 \end{array}\right\} \Rightarrow t = 2{,}262$$

d'où :

$$P\left[20 - \left(2{,}262 \times \frac{2{,}91}{\sqrt{10}}\right) < \bar{x} \leq 20 + \left(2{,}262 \times \frac{2{,}91}{\sqrt{10}}\right)\right] = 0{,}95$$

$$P(17{,}92 < \bar{x} \leq 22{,}08) = 0{,}95$$

4°) *Commentaire*

a) avec $n = 100$, $\bar{x} \to N\left(m, \frac{\sigma}{\sqrt{n}}\right)$ sans avoir d'autres précisions

avec $n = 25$: il faut connaître la loi du caractère dans la population pour déduire celle de $\bar{x}$. Il est dit que X suit une loi normale. Par conséquent

$$\bar{x} \to N\left(m, \frac{\sigma}{\sqrt{n}}\right)$$

avec $n = 10$, σ^2 est inconnue.

On l'estime par s^2. Alors $\bar{x}$ suit une loi de Student-Fisher à $(n - 1)$ degrés de liberté.

b) L'étendue de l'intervalle s'aggrandit avec la diminution de la taille de l'échantillon, c'est-à-dire que la précision de l'encadrement de $\bar{x}$ diminue au fur et mesure que n diminue. Pour que l'échantillon soit représentatif, il faut qu'il ait une taille suffisamment grande.

EXERCICE N° 115

Dans une étude de marchés, un calcul économique a montré que le lancement d'un produit n'est rentable que si 20 % des individus sont favorables à la consommation de ce produit.

On fait des sondages en étudiant des échantillons de 100 consommateurs.

1°) Calculer la probabilité pour que les proportions de personnes favorables soient comprises entre 15 % et 20 %.

2°) Un sondage sur 1 600 consommateurs a montré que 192 personnes étaient favorables. Si l'on accepte un risque d'erreur de 2,5 % quelle décision doit prendre la direction de la firme ?

(Licence Sc. Eco. 2e année Juin 19[illegible])

Solution

1°) Probabilité que f soit comprise entre 15 % et 20 % :

Dans l'exercice n° 100 (Loi des grands nombres) nous avons montré que f suit une loi normale de paramètres :

$$E(f) = p = 0{,}20$$

$$\sigma(f) = \sqrt{\frac{pq}{n}} = 0{,}04$$

Par conséquent

$$P\left\{p - t\sqrt{\frac{pq}{n}} < f \le p + t\sqrt{\frac{pq}{n}}\right\} = \int_{-t}^{+t} \frac{1}{\sqrt{2\Pi}} e^{-\frac{t^2}{2}} dt$$

$$= 2\,\Pi(t) - 1$$

la borne inférieure est :

$$0{,}20 - t_1 \sqrt{\frac{0{,}20 \times 0{,}80}{100}} = 0{,}15 \Rightarrow t_1 = -1{,}25$$

... la borne supérieure est :

$$0{,}20 + t_2 \sqrt{\frac{0{,}20 \times 0{,}80}{100}} = 0{,}25 \Rightarrow t_2 = +1{,}25$$

d'où la probabilité :

$2\Pi(1{,}25) - 1 = 2 \times 0{,}8944 - 1 = 0{,}7888$

Il y a 78,88 % de chances que la proportion de personnes favorables au produit dans la population soit comprise entre 15 et 25 pour cent.

2°) Décision à prendre :

Il convient au préalable d'établir l'encadrement de la proportion théorique de personnes favorables dans un intervalle à 97,5 % de chances, puisqu'on accepte un risque de 2,5 %.

$$P\left\{f - t\sqrt{\frac{f(1-f)}{n}} < p \le f + t\sqrt{\frac{f(1-f)}{n}}\right\} = 0{,}975$$

$$P\left\{\frac{192}{1\,600} - 2{,}248\sqrt{\frac{0{,}12 \times 0{,}88}{1\,600}} < p \le \frac{192}{1\,600} + 2{,}248\sqrt{\frac{0{,}12 \times 0{,}88}{1\,600}}\right\} = 0{,}975$$

car $2\Pi(t) - 1 = 0{,}975 \Rightarrow \Pi(t) = 0{,}9875 \Rightarrow t = 2{,}248$

d'où

$P\{0{,}102 < p \le 0{,}138\} = 0{,}975$

or l'hypothèse faite d'avoir 20 % de gens favorables dans la population est incompatible (p n'étant pas compris dans l'intervalle calculé) dans la limite du risque admis. Par conséquent le directeur de la firme ne doit pas lancer le produit

EXERCICE N° 116

Une machine de fabrication en séries produit des pièces dont le diamètre est une variable aléatoire X normale, de moyenne 32 mm et d'écart-type 1 mm. Pour contrôler la fabrication, on prélève à intervalles réguliers 20 pièces ; soit $\bar{x}$ la moyenne des diamètres dans un échantillon.

1. quelle est la loi de probabilité de $\bar{x}$? Justifier votre réponse

2. en rappelant brièvement le principe d'utilisation d'une carte de contrôle, donner les limites (a, b) qui doivent contenir $\bar{x}$ pour que la machine puisse être considérée comme bien réglée avec une probabilité de 0,99 ?

3. pour être acceptées et utilisées, les pièces doivent satisfaire à la norme suivante $31 < X < 33$.
   Quelle est la probabilité pour qu'une pièce soit utilisable ?

4. le coût de fabrication d'une pièce est 10,80 DH. Pour diminuer le pourcentage de pièces défectueuses, on peut utiliser une machine plus moderne ; l'écart-type de X sera alors de 0,5 mm, mais le prix de revient est de 12,00 DH.
   En comparant le coût réel d'une *bonne* pièce, dire si on a intérêt à choisir la nouvelle machine.

(Licence Sc. Eco. 2e année - Juin 197[illegible])

Solution

1. *Loi de $\bar{x}$* :

Le diamètre X est une variable aléatoire normale, par conséquent $Y = \sum X$ suit aussi une loi normale (théorème-central-limite)

d'où :

$$t = \frac{Y - E(Y)}{\sigma_Y} \rightarrow N(0, 1)$$

ou encore :

$$t = \frac{\sum X_i - E[\sum X_i]}{\sigma[\sum X_i]} \rightarrow N(0, 1)$$

or on sait que :

$E[\sum X_i] = \sum E(X_i) = n \cdot m$

$V(\sum X_i) = \sum V(X_i) = n\sigma^2$

$\sigma(\sum X_i) = \sigma \cdot \sqrt{n}$

D'où :

$$t = \frac{\sum X_i - n \cdot m}{\sigma\sqrt{n}} \rightarrow N(0,1)$$

on divise l'ensemble des termes par n :

$$t = \frac{\frac{\Sigma X_i}{n} - \frac{n\,m}{n}}{\frac{\sigma\sqrt{n}}{n}} \rightsquigarrow N(0,1)$$

$$t = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$$

Or on sait que $E(\bar{x}) = m$ et $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Il en résulte que

$$\bar{x} \rightsquigarrow N\left(m ; \frac{\sigma}{\sqrt{n}}\right)$$

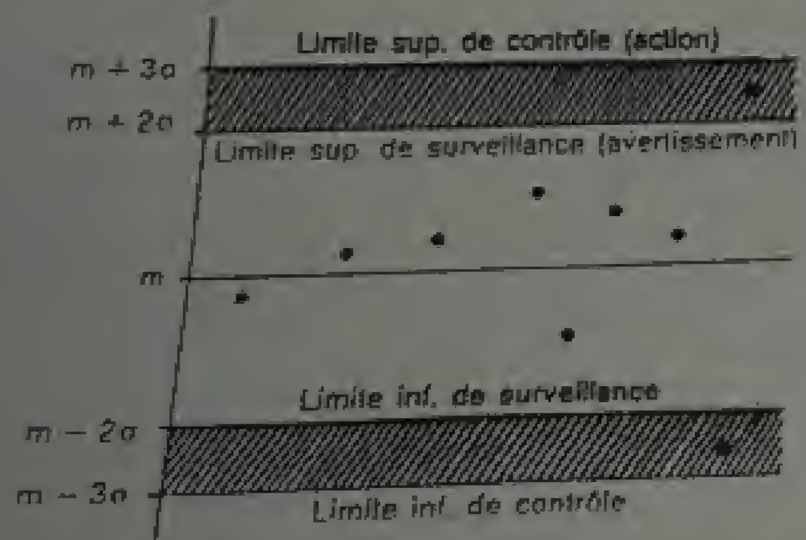
2. Principe d'utilisation de la carte de contrôle :

La carte de contrôle est un procédé qui permet de contrôler si un processus de fabrication demeure valable et d'alerter le chef de production de toute défaillance intervenue dans le réglage de la machine.

Elle comprend deux intervalles :

— l'une, à 95 % de chances, dont les bornes constituent les limites de surveillance.

— l'autre, à 99,7 % de chance, dont les bornes constituent les limites de contrôle.



Pour contrôler la moyenne des articles fabriqués on prélève à intervalle de temps régulier un échantillon de n articles, on détermine leur moyenne. On représente cette moyenne sur la carte.

Si le point représentatif se trouve en dehors des limites de contrôle (phénomène ayant une probabilité très faible de se produire ; 0,3 %), on peut affirmer avec peu de chances de se tromper que la machine est déréglée : et on doit arrêter la fabrication.

Si le point représentatif de la moyenne de l'échantillon extrait se trouve à l'intérieur des limites de surveillance, on peut affirmer que la machine est toujours réglée.

Si, enfin, le point représentatif tombe dans la zone hachurée (entre les deux limites), on prélève immédiatement un 2e échantillon ; si le point représentatif est toujours dans cette zone il faut stopper la fabrication pour procéder au réglage de la machine.

Ceci étant étant rappelé, calculons l'intervalle (a, b) à 99 % :

$$P\left(m - t\frac{\sigma}{\sqrt{n}} < \bar{x} \leq m + t\frac{\sigma}{\sqrt{n}}\right) = 2\Pi(t) - 1 = 0,99$$

$2\Pi(t) - 1 = 0,99 \Rightarrow t = 2,58$ (table de la fonction intégrale de la loi de Laplace-Gauss)

$$P\left(32 - 2,58 \cdot \frac{1}{\sqrt{20}} < \bar{x} < 32 + 2,58 \cdot \frac{1}{\sqrt{20}}\right) = 0,99$$

$$P\left(32 - \frac{2,58}{4,47} < \bar{x} \leq 32 + \frac{2,58}{4,47}\right) = 0,99$$

soit

$$P(31,42 < \bar{x} \leq 32,58) = 0,99$$

3. $P(31 < X < 33) = \Pi\left(\frac{33 - 32}{1}\right) - \Pi\left(\frac{31 - 32}{1}\right)$

$$= \Pi(1) - \Pi(-1)$$
$$= 2\Pi(1) - 1$$
$$= 2 \times 0,8413 = 0,6826$$

Il y a 68 % de chances qu'une pièce soit utilisable ; soit sur 100 pièces, 68 sont valables.

4. Il s'agit de comparer le coût réel d'une pièce valable :

— Pour la 1re machine : $10,80 \times \frac{100}{68} = 15,88$ DH

— Pour la deuxième machine calculons d'abord le taux de bonnes pièces

$$P(31 < X < 33) = \Pi\left[\frac{33 - 32}{\frac{1}{2}}\right] - \Pi\left[\frac{31 - 32}{\frac{1}{2}}\right]$$

$$= \Pi(2) - \Pi(-2)$$
$$= 2\,\Pi(2) - 1$$
$$= 2 \times 0,9772 - 1 = 0,9544$$

D'où le coût réel d'une pièce valable est :

$$12 \times \frac{100}{95} = 12,63$$

Donc on a intérêt à changer de machine.

INTRODUCTION À LA THÉORIE DE L'ESTIMATION

Des observations nombreuses ont conduit à admettre que, dans une population donnée, la fréquence des individus allergiques à une substance donnée est 0,2.

I. On tire un échantillon de taille 100 dans cette population :

1°) Quelle est la probabilité d'observer k personnes allergiques dans l'échantillon ? Quel est le nombre le plus probable ?

2°) L'examen de l'échantillon montre que k = 25. Ce résultat est-il compatible au seuil 0,05 de rejet avec l'hypothèse faite au départ ?

3°) A partir du seul résultat de l'échantillon, encadrer à 95 % la fréquence exacte des individus allergiques dans la population.

II. On envisage maintenant le cas où l'échantillon a pour taille 400 et fournit, après examen, 100 personnes allergiques.

Reprendre les questions de I.

III. Déduire de ce qui précède qu'une affirmation du genre « un échantillon a fourni la fréquence 0,25 » est insuffisante en elle-même et calculer à partir de quelle taille de l'échantillon elle devient significative avec une probabilité de 0,95.

Solution

I. *La taille de l'échantillon n = 100 :*

1°) Soit X « le nombre de personnes allergiques dans l'échantillon »

$P(X = k) = C_{100}^{k}\,(0.2)^k\,(0.8)^{100-k}$

X étant une variable binomiale qui peut être approchée par une loi normale car n est grand et p n'est voisin ni de 0 ni de 1 ; ses paramètres sont :

$E(X) = np = 20$

$\sigma_x = \sqrt{npq} = 4$

La valeur la plus probable est comprise ou égale à np − q et np + p or :

$np - q = 20 - 0.80 = 19.20$

$np + p = 20 + 0.20 = 20.20$

Le mode est donc égal à 20, nombre entier compris entre 19,20 et 20,20.

2°) L'échantillon fournit 25 personnes allergiques

$k = 25$ et $f = 0{,}25$

Par ailleurs on sait que

$$f \to N\left(p = 0{,}20\ ;\ \sqrt{\frac{pq}{n}} = 0{,}04\right)$$

d'où

$$P\left\{p - t\sqrt{\frac{pq}{n}} < f \leq p + t\sqrt{\frac{pq}{n}}\right\} = 2\pi(t) - 1 = 0{,}95$$

$$P\{0{,}20 - 2 \times 0{,}04 < f \leq 0{,}20 + 2 \times 0{,}04\} = 0{,}95$$

t étant égal à 1,96, soit 2 :

$$P\{0{,}12 < f \leq 0{,}28\} = 0{,}95$$

f est bien contenue dans l'intervalle ainsi calculé, par conséquent le résultat fourni par l'échantillon est compatible avec le résultat $p = 0{,}20$

3°) Intervalle contenant p au seuil 0,05 :

$$P\left\{f - t\sqrt{\frac{f(1-f)}{n}} < p \leq f + t\sqrt{\frac{f(1-f)}{n}}\right\} = 2\pi(t) - 1 = 0{,}95$$

$$P\left\{0{,}25 - 2\sqrt{\frac{0{,}25 \cdot 0{,}75}{100}} < p \leq 0{,}25 + 2\sqrt{\frac{0{,}25 \cdot 0{,}75}{100}}\right\} = 2\pi(t) - 1$$

$$P\{0{,}17 < p \leq 0{,}33\} = 0{,}95$$

L'intervalle trouvé contient bien $p = 0{,}20$

II *La taille de l'échantillon est 400 et les personnes allergiques au nombre de 100*

1°) La variable X suit toujours une loi binomiale de paramètres 400 et 0,2

$X \to \mathcal{B}(400\ ;\ 0{,}2)$

L'approximation par une loi normale est évidente ; le nombre le plus probable est cette fois égal à 80

2°) $$P\left\{0{,}2 - 2\sqrt{\frac{0{,}2 \times 0{,}8}{400}} < f \leq 0{,}2 + 2\sqrt{\frac{0{,}2 \times 0{,}8}{400}}\right\} = 2\pi(t) - 1 = 0{,}95$$

$$P\{0{,}16 < f \leq 0{,}24\} = 0{,}95$$

Le résultat fourni par l'échantillon est incompatible avec l'hypothèse $p = 0{,}20$

3°) $$P\left\{f - t\sqrt{\frac{f(1-f)}{n}} < p \leq f + t\sqrt{\frac{f(1-f)}{n}}\right\} = 0{,}95$$

$$P\left\{0{,}25 - 2\sqrt{\frac{0{,}25 \times 0{,}75}{400}} < p \leq 0{,}25 + 2\sqrt{\frac{0{,}25 \times 0{,}75}{400}}\right\} = 0{,}95$$

$$P\{0{,}21 < p \leq 0{,}29\} = 0{,}95$$

ce qui confirme le rejet de l'hypothèse que p est égale à 0,20.

III *Détermination de n*

On constate que

Avec $n = 100$ aucune incompatibilité entre l'hypothèse faite sur p et le résultat de l'échantillon.

Avec $n = 400$ au contraire $p = 0{,}20$ est incompatible

Entre 100 et 400 il y a une taille optimale qu'il faut préciser pour que $p = 0{,}20$ cesse d'être compatible

Soit n cette taille ; elle entraînera le rejet de l'hypothèse faite sur p si

$$0{,}25 > 0{,}2 + 2\sqrt{\frac{0{,}2 \times 0{,}8}{n}}$$

$$0{,}25 > 0{,}2 + \frac{2 \times 0{,}4}{\sqrt{n}}$$

$$0{,}05 > \frac{0{,}8}{\sqrt{n}} \Rightarrow \sqrt{n} > \frac{80}{5} \Rightarrow n > (16)^2$$

La taille de l'échantillon doit dépasser 256

EXERCICE N° 118

Un sondage relatif à la consommation de cigarettes est effectué auprès de 100 personnes d'une ville déterminée. L'échantillon est supposé représentatif de l'ensemble de la population de la ville. Les résultats du sondage sont les suivants

Nombre de personnes qui fument	Nombre de paquets de cigarettes fumées par jour
n_i	x_i
2	0
34	1
40	2
13	3
8	4
3	5

1°) Encadrer la proportion de personnes qui fument dans la population avec une probabilité égale à 0,95. En déduire entre quels chiffres varie le nombre de fumeurs pour une ville de 20 000 habitants ?

2°) Calculer la consommation moyenne dans l'échantillon désignée par $\bar{x}$. Encadrer la consommation moyenne de cigarettes dans la population appelée m avec une probabilité égale à 0,95

On donne $$\sum_{i=1}^{6} n_i (x_i - \bar{x})^2 = 114$$

3°) Déterminer la taille de l'échantillon à interroger pour connaître la consommation moyenne journalière de paquets de cigarettes dans la population avec une erreur absolue 0,1 $\left(\frac{1}{10}\text{ de paquet}\right)$

(On considère toujours que la probabilité de l'intervalle est de 0,95)

(Licence 2e année - Sciences économiques - Juin 1975)

Solution

1°) Intervalle de confiance à 95 % de chances pour p :

$$P\left[f - t\sqrt{\frac{f(1-f)}{n}} < p \leq f + t\sqrt{\frac{f(1-f)}{n}}\right] = 2\pi(t) - 1 = 0{,}95$$

$$f = \frac{\text{nombre de personnes qui fument}}{\text{taille de l'échantillon}} = \frac{100-2}{100} = 0{,}98$$

$2\pi(t) - 1 = 0{,}95 \Rightarrow t = 1{,}96 \# 2$

d'où la borne inférieure : $0{,}980 - 2\sqrt{\frac{0{,}98 \times 0{,}02}{100}} = 0{,}952$

et la borne supérieure : $0{,}980 + 0{,}028 \# 1$

$P(0{,}952 < p \leq 1) = 0{,}95$

Pour une population de 20 000 habitants le nombre de fumeurs est :

au minimum $20\,000 \times 0{,}952 = 19\,040$

au maximum $20\,000 \times 1 = 20\,000$

Il y a donc 95 % de chances pour que le nombre de fumeurs varie entre 19 040 et 20 000.

2°) Intervalle de confiance à 95 % pour $\bar{x}$:

n_i	x_i	$n_i x_i$
2	0	0
34	1	34
40	2	80
13	3	39
8	4	32
3	5	15
		200

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{200}{100} = 2$$

L'intervalle de confiance de m est telle que

$$P[\bar{x} - t\sigma_{\bar{x}} < m \leq \bar{x} + t\sigma_{\bar{x}}] = \int_{-t}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 2\pi(t) - 1$$

Pour calculer $\sigma_{\bar{x}}$, on sait que $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

σ étant l'écart-type de la population qu'il faut estimer. Or la taille de l'échantillon étant grande, la variance de la population peut être approchée valablement par celle de l'échantillon, celle-ci étant donnée par

$$\sigma_e^2 = \frac{\sum_{i=1}^{6} n_i(x_i - \bar{x})^2}{n} = \frac{114}{100} = 1{,}14$$

d'où $\sigma = \sigma_e = \sqrt{1{,}14} = 1{,}0678$

$$\sigma_{\bar{x}} = \frac{1{,}0678}{10} = 0{,}107$$

Remarque L'estimation de σ^2 est en fait fournie par $s^2 = \frac{\sum n_i(x_i - \bar{x})^2}{n-1}$

soit $\frac{114}{99} = 1{,}15$ d'où $\sigma \# s = 1{,}07$

Mais n étant grand, le biais entre σ^2 et σ_e^2 est négligeable

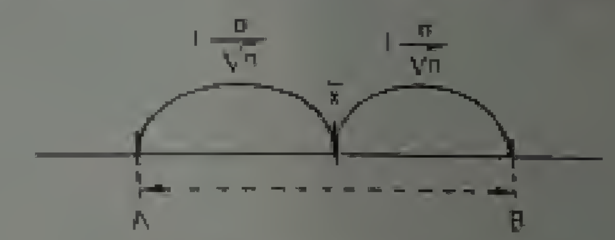
Quoiqu'il en soit l'intervalle de confiance de m est

borne inférieure : $2 - 2 \times 0{,}107 = 1{,}786$

borne supérieure : $2 + 2 \times 0{,}107 = 2{,}214$

Il y a donc 95 % de chances pour que la consommation moyenne de cigarettes par jour soit comprise entre 1,8 et 2,2 paquets, soit 36 et 44 cigarettes.

3°) *Calcul de n.*



[AB] = Intervalle de confiance contenant m

Ainsi l'écart maximum entre $\bar{x}$ et m est égal à $t\frac{\sigma}{\sqrt{n}}$. Il s'agit de déterminer n de sorte que cet écart ne dépasse pas 0,1 avec une probabilité de 0,95. D'où :

$$t\frac{\sigma}{\sqrt{n}} \leq 0{,}1 \Rightarrow n \geq \frac{t^2\sigma^2}{10^{-2}}$$

$\alpha = 0{,}95 \Rightarrow t = 1{,}96 \# 2$ et $\sigma^2 = 1{,}14$ par conséquent

$n \geq 4 \times \frac{1{,}14}{10^{-2}}$ n doit être supérieur à 456.

EXERCICE N° 119

Quelques jours avant les consultations électorales auxquelles se présentent deux candidats A et B, un organisme procède à un sondage aléatoire dans l'ensemble de la population pour estimer le pourcentage de voix qu'obtiendront les deux candidats.

Soit P le pourcentage de voix du candidat A.

1°) Quel doit être le nombre d'individus interrogés pour que l'amplitude de l'intervalle de confiance à 0,95 ne soit pas supérieure à 0,02 ? On traitera cette question en supposant successivement que :

a) $P \leq 30\,\%$

b) l'on n'a aucune idée a priori sur P.

2°) On interroge 200 individus, 80 se prononcent pour A.

Donner une estimation du pourcentage des voix obtenues par A ainsi que la précision de cette estimation avec une probabilité de 95 %.

Solution

1°) Calcul du nombre d'individus à interroger :

Soit f la proportion des individus favorables au candidat A dans l'échantillon de taille n à déterminer. On sait que

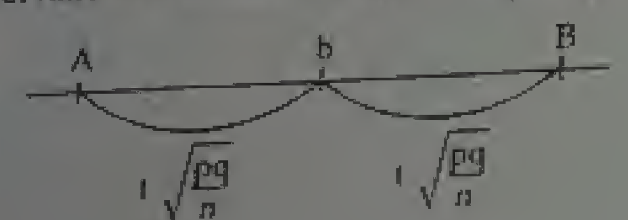
$$f \to N\left(p ; \sqrt{\frac{pq}{n}}\right)$$

Il en résulte que

$$P\left[p - t\sqrt{\frac{p(1-p)}{n}} < f \leq p + t\sqrt{\frac{p(1-p)}{n}}\right] = 2\pi(t) - 1 = 0,95$$

$$\alpha = 0,95 \Rightarrow t = 1,96 \simeq 2$$

L'amplitude de l'intervalle de confiance est égale à $2t\sqrt{\frac{p(1-p)}{n}}$



d'où

$$4\sqrt{\frac{p(1-p)}{n}} < 0,02 \Rightarrow \frac{16\,p(1-p)}{n} < 4 \cdot 10^{-4}$$

$$\frac{16\,p(1-p)}{4 \cdot 10^{-4}} < n$$

ou bien $n > 4 \cdot 10^4\,p(1-p)$

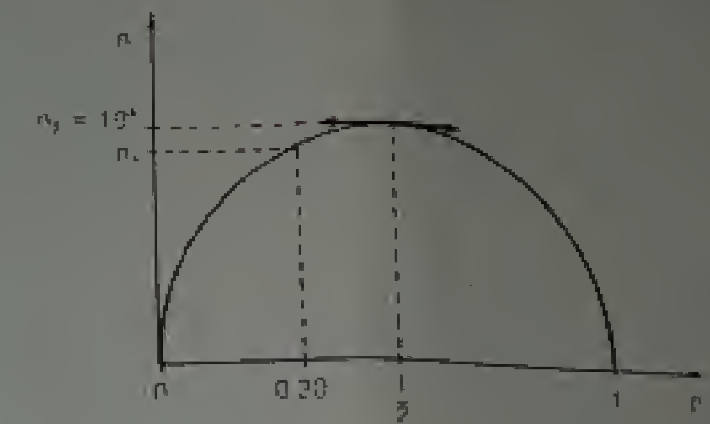
A la limite n est une fonction du second degré de p

$$n = -4 \cdot 10^4 p^2 + 4 \cdot 10^4 p$$

Calculons sa dérivée

$$n' = -8 \cdot 10^4 p + 4 \cdot 10^4 = 4 \cdot 10^4(-2p + 1)$$

Elle s'annule pour la valeur $p = \frac{1}{2}$. D'où le graphe



Ainsi :

a) si $p = 30\,\% \Rightarrow n_1 = 4 \cdot 10^4 \times 0,3 \times 0,7 = 8\,400$

b) si p est inconnue on prendra la taille maximale possible qui est donnée pour $p = \frac{1}{2}$

d'où $n_2 = 4 \cdot 10^4 \times \frac{1}{2} \times \frac{1}{2} = 10\,000$

2°) Estimation ponctuelle et par intervalle de confiance de p

Une bonne estimation de p est fournie par f car :

$E(f) = p$ (absence de biais)

$V(f) = \frac{pq}{n} \to 0$ si $n \to \infty$ (Convergence)

ainsi f étant égale à $\frac{80}{200} = 0,40$, il y aurait 40 % d'individus favorables à A dans l'ensemble de la population.

La précision de cette estimation par un ensemble de confiance à 95 % est alors

$$P\left[f - 2\sqrt{\frac{f(1-f)}{n}} < p \leq f + 2\sqrt{\frac{f(1-f)}{n}}\right] = 0,95$$

$$\alpha = 0,95 \Rightarrow t = 1,96 \simeq 2$$

$$P\left[0,40 - 2\sqrt{\frac{0,4 \times 0,6}{200}} < p \leq 0,40 + 2\sqrt{\frac{0,4 \times 0,6}{200}}\right] = 0,95$$

$$P\,(0,33 < p \leq 0,47) = 0,95$$

Il y a 95 % de chances que le pourcentage de voix favorables au candidat A soit compris entre 33 et 47 pour cent.

EXERCICE N° 120

Un chef d'entreprise occupe 10 000 employés. Il se propose d'aménager un emplacement pour que ses employés puissent y garer leurs voitures.

1°) Il sait que le nombre de voitures dépasse 2 000 ; quelle doit être la taille de l'échantillon formé d'employés qu'il se propose d'interroger, s'il veut connaître le résultat à 10 % près avec une probabilité de 95 %.

2°) Il décide d'interroger 1 600 employés ; de leurs réponses, il résulte que 640 viennent avec leur voiture. Estimer le nombre de places à aménager, ponctuellement et par un ensemble de confiance au seuil 0,05. Est-ce que le seuil 0,05 de rejet signifie, à votre avis, que la capacité des emplacements se révèlera insuffisante moins de 5 jours sur 100, soit environ 18 jours ?

Solution

1°) Taille de l'échantillon à interroger :

Il s'agit de trouver la relation entre la taille de l'échantillon et la précision relative à l'estimation, précision qui est égale ici à 10 % c'est-à-dire $\frac{1}{10}$; à partir de l'intervalle de confiance, l'écart entre f et p est au maximum égal à

$$t\sqrt{\frac{pq}{n}}$$

L'erreur relative commise est alors :

$$\frac{t\sqrt{\frac{pq}{n}}}{p}$$

celle-ci ne doit pas dépasser 10 % ; d'où :

$$\frac{t\sqrt{\frac{pq}{n}}}{p} \leq \frac{1}{10}$$

or $\alpha = 0{,}95 \Rightarrow t \neq 2$ et $p = \frac{2\,000}{10\,000} = 0{,}20$

$$\frac{2\sqrt{\frac{p(1-p)}{n}}}{p} \leq \frac{1}{10} \Rightarrow \frac{p(1-p)}{n} \leq \frac{p^2}{400} \Rightarrow n \geq 400 \cdot \frac{1-p}{p}$$

$n \geq 400 \times \frac{1-0{,}2}{0{,}2}$ soit *1600*.

si l'on voudrait connaître la valeur de p à 10 % avec une probabilité de 0,95, il faudrait interroger au moins 1 600 individus.

2°) Estimation du nombre de places à aménager :

L'échantillon de 1 600 personnes a montré que 640 viennent avec leur voiture soit une proportion :

$$f = \frac{640}{1\,600} = 0{,}40$$

L'estimation ponctuelle de p est valablement fournie par f soit 40 % d'où le nombre de places à prévoir serait de :

$0{,}40 \times 10\,000 = 4\,000$

Intervalle de confiance à 0,95 pour p :

On encadre la proportion exacte p des utilisateurs de voitures par

$$f - t\sqrt{\frac{f(1-f)}{n}} < p \leq f + t\sqrt{\frac{f(1-f)}{n}}$$

or $\alpha = 0{,}95 \Rightarrow t \neq 2$ et $f(1-f) = 0{,}24 \neq \frac{1}{4}$

$$2\sqrt{f(1-f)} = 2\sqrt{\frac{1}{4}} = 1$$

d'où

$$0{,}4 - \frac{1}{40} < p \leq 0{,}4 + \frac{1}{40}$$

$$P(0{,}375 < p \leq 0{,}425) = 0{,}95$$

Ainsi pour un intervalle de confiance à 0,95, il suffira d'aménager 4 250 places (10 000 × 0,425).

Il ne faut pas se tromper sur l'interprétation du seuil de rejet. En effet, la borne supérieure s'écarte de p (estimée ponctuellement) de 0,025 ; On serait alors tenté de penser que l'aménagement sera insuffisant 9 jours (2,5 % × 360 jours) par an. En fait, il résulte de la théorie de l'estimation que p a une valeur inconnue mais certaine et que l'encadrement de p à 0,95 d'être sûr donne une fourchette susceptible de contenir la vraie valeur de p avec un risque d'erreur de 5 %.

EXERCICE N° 121

Un candidat à une élection veut connaître la proportion de voix qui lui sont favorables en procédant par sondage sur la population.

Un échantillon de 100 électeurs donne une fréquence observée f = 0,55 en sa faveur.

1°) Quel est l'intervalle de confiance pour la proportion p d'électeurs favorables à ce candidat dans la population, au niveau 0,95 fourni par l'abaque ?

2°) Quel est l'intervalle de confiance obtenu par le calcul direct pour le même coefficient de confiance ?

Solution

1°) *Lecture de l'abaque*

p sera lu en ordonnée pour une abscisse 0,55 et les deux branches correspondant à $n = 100$

On relève

— branche inférieure : $p_1 = 45\ \%$
— branche supérieure : $p_2 = 65\ \%$

Il y a 95 % de chances pour que la proportion d'électeurs favorables dans le corps électoral soit comprise entre 45 % et 65 %

2°) *Intervalle de confiance :*

borne supérieure : $f + t\sqrt{\frac{pq}{n}}$

borne inférieure : $f - t\sqrt{\frac{pq}{n}}$

p sera approchée par f et $t \approx 2$ ($\alpha = 0{,}95$), d'où :

borne supérieure : $0{,}55 + 2\sqrt{\dfrac{0{,}55 \times 0{,}45}{100}} = 0{,}6495$

borne inférieure : $0{,}55 - 2\sqrt{\dfrac{0{,}55 \times 0{,}45}{100}} = 0{,}4505$

Les résultats fournis par l'abaque et le calcul sont sensiblement identiques.

EXERCICE N° 122

On veut estimer la consommation moyenne des ménages dont les revenus annuels sont compris entre 15 000 et 25 000 Dirham.

1°) Quelle est la différence entre un estimateur sans biais et un estimateur asymptotiquement sans biais ? Quelles sont les caractéristiques d'un estimateur correct ? Illustrer vos réponses.

2°) Un échantillon de 10 ménages a donné les résultats suivants :

$$\sum_{i=1}^{10} x_i = 215\,980 \text{ DH}$$

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 55\,164\,110 \text{ DH}$$

où x_i désigne la consommation du ménage i

Donner une estimation de la consommation moyenne des ménages dont les revenus annuels sont compris entre 15 000 et 25 000 DH. Préciser cette estimation dans un intervalle de confiance à 0,95

3°) Supposons maintenant que l'on veuille déterminer la proportion des ménages dont la consommation se situe entre 19 800 et 21 600 DH. Si l'on désire obtenir, en tirant un échantillon, une estimation qui soit, avec une probabilité 0,9 à moins de 0,03 unité de la valeur correcte, quelle doit être la taille de l'échantillon étant donné :

a) que l'on sait la proportion réelle proche de 35 % ?
b) que l'on n'a aucune idée de la proportion réelle ?

(Licence 2e année - Sciences économiques [illegible])

Solution

1°) *Notion d'estimateur :*

Un estimateur $\hat{\varphi}$ est dit sans biais si
$E(\hat{\varphi}) = \varphi$
Il est asymptotiquement sans biais si :
$E(\hat{\varphi}) \to \varphi \quad \text{si } n \to \infty$

L'estimateur est convergent si
$V(\hat{\varphi}) \to 0 \quad \text{quand } n \to \infty$

Un estimateur est correct s'il est sans biais et convergent

Exemples :

a) σ_e^2 est un estimateur biaisé de σ^2 car

$$E(\sigma_e^2) = \left(1 - \frac{1}{n}\right)\sigma^2$$

$$\text{si } n \to \infty : 1 - \frac{1}{n} \to 1 \Rightarrow E(\sigma_e^2) \to \sigma^2$$

Le biais disparaît si la taille de l'échantillon est grande

b) $S^2 = \dfrac{n}{n-1}\sigma_e^2$ est un estimateur correct de σ^2 car

$$E(S^2) = \frac{n}{n-1}E(\sigma_e^2) = \frac{n}{n-1} \times \frac{n-1}{n}\sigma^2 = \sigma^2$$

2°) *Estimation de la consommation moyenne :*

Soit $\bar{x}$ la consommation moyenne dans l'échantillon :

$$\bar{x} = \frac{215\,980}{10} = 21\,598$$

$\bar{x}$ donne une estimation correcte de m ; ainsi la consommation moyenne de la population serait de 21 598 DH

Intervalle de confiance :

$$P\left\{\bar{x} - t\frac{\sigma}{\sqrt{n}} < m \leq \bar{x} + t\frac{\sigma}{\sqrt{n}}\right\} = 0{,}95$$

La variance σ^2 de la population étant inconnue sera estimée par s^2 car l'échantillon est de petite taille, s est une variable telle que :

$$t = \frac{\bar{x} - m}{s/\sqrt{n}}$$

suit une loi de STUDENT-FISHER à 9 degrés de liberté.

L'intervalle de confiance de m est alors :

$$P\left\{\bar{x} - t\frac{s}{\sqrt{n}} < m \leq \bar{x} + t\frac{s}{\sqrt{n}}\right\} = 0{,}95$$

$$s^2 = \frac{55\,164\,110}{9} = 6\,129\,345$$

$$s \simeq 2\,475{,}75$$

$$\left.\begin{array}{l}\alpha = 0{,}95 \Rightarrow 1 - \alpha = 0{,}05 \\ V \quad = n - 1 = 9\end{array}\right\} \Rightarrow t = 2{,}262$$

borne inférieure : $21\,598 - \left(\frac{2{,}262 \times 2\,475{,}75}{\sqrt{10}}\right)$

$$= 21\,598 - 1\,772 = 19\,826$$

borne supérieure : $21\,598 + \left(\frac{2{,}262 \times 2\,475{,}75}{\sqrt{10}}\right)$

$$= 21\,598 + 1\,772 = 23\,370$$

La consommation moyenne serait comprise entre 19 826 et 23 370 avec une probabilité de 95 pour cent.

3°) *Détermination de la taille de l'échantillon*

$$\alpha = 0{,}99 \Rightarrow t = 2{,}58$$

$$t\sqrt{\frac{p(1-p)}{n}} \leq 0{,}03$$

$$t^2\frac{p(1-p)}{n} \leq 9 \cdot 10^{-4} \Rightarrow \frac{t^2 p(1-p)}{9 \cdot 10^{-4}} \leq n$$

$$n \geq \frac{(2{,}58)^2}{9} \cdot 10^4 \times p(1-p)$$

$$n \geq \left(\frac{2{,}58}{3}\right)^2 \cdot 10^4 \times p(1-p)$$

$$n \geq (0{,}86)^2 \cdot 10^4 \times p(1-p)$$

$$n \geq 0{,}7396 \cdot 10^4 \times p(1-p)$$

$$n \geq 7\,400\, p(1-p)$$

a) $p = 0{,}35$:

$$n \geq 7\,400 \times 0{,}35 \times 0{,}65$$

$$n \geq 7\,400 \times 0{,}2\,275$$

n doit être supérieur à 1 684

b) p inconnu

On prendra le maximum possible obtenu pour $p = q = \frac{1}{2}$

(cf exercice n° 119)

$$n \geq 7\,400 \times \frac{1}{2} \times \frac{1}{2}$$

n doit être supérieur à 1 850

EXERCICE N° 123

On estime la variance d'une population par la quantité

$$\frac{\sum_{i=1}^{n} (x_i - m)^2}{n}$$

où x_i représente la variable, i la valeur observée dans l'échantillon extrait avec remise et m la moyenne du caractère dans la population. Dire si la quantité ci-dessus représente une estimation sans biais de σ^2 variance de la population

(Licence 2° année Sciences économiques - Juin 1975)

Solution

La quantité

$$\frac{\sum_{i=1}^{n} (x_i - m)^2}{n}$$

représentera un estimateur sans biais de σ^2 si son espérance est égale à σ^2

$$E\left[\frac{\sum_{i=1}^{n} (x_i - m)^2}{n}\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} (x_i - m)^2\right] \qquad = \frac{1}{n}\sum_{i=1}^{n} E[(x_i - m)^2]$$

or

$E(x_i - m)^2 = \sigma^2$ (par définition)

Par conséquent

$$E\left[\frac{\sum_{i=1}^{n}(x_i - m)^2}{n}\right] = \sigma^2$$

$$\frac{\sum_{i=1}^{n}(x_i - m)^2}{n}$$ est un estimateur sans biais de σ^2

EXERCICE N° 124

Un sondage aléatoire portant sur un échantillon de 2 100 électeurs a donné 630 individus favorables au candidat A.

Calculer l'intervalle de confiance à 1 % de risque pour la proportion des électeurs favorables à A dans la population totale des électeurs.

(Licence 2ᵉ année Sciences économiques Octobre 1977)

Solution

Soit X « le nombre des électeurs favorables à A ».

X est une variable binomiale de paramètres :

$n = 2\,100$

$$p = \frac{630}{2\,100} = \frac{3}{10} = 0{,}3$$

$$X \to \mathcal{B}\left(2\,100 \,;\, \frac{3}{10}\right)$$

Par ailleurs n est grand et p est loin de 0 et 1 ; la loi binomiale peut être approchée par une loi normale de paramètres :

np et $\sqrt{npq}$

alors

$$t = \frac{X - np}{\sqrt{npq}} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} = \frac{f - E(f)}{\sigma_f}$$

suit une loi normale centrée et réduite

Il en résulte que

$$f \to N\left(p \,;\, \sqrt{\frac{pq}{n}}\right)$$

d'où

$$P\left[\frac{3}{10} - t\sqrt{\frac{f(1-f)}{n}} < p \leq \frac{3}{10} + t\sqrt{\frac{f(1-f)}{n}}\right] = 2\pi(t) - 1 = 0{,}99$$

$2\pi(t) = 1{,}99 \quad t = 2{,}58 \approx 2{,}6$

$$P\left[0{,}3 - 2{,}6\sqrt{\frac{0{,}3 \times 0{,}7}{2\,100}} < p \leq 0{,}3 + 2{,}6\sqrt{\frac{0{,}3 \times 0{,}7}{2\,100}}\right] = 0{,}99$$

$$P\left[0{,}3 - \frac{2{,}6}{100} < p \leq 0{,}3 + \frac{2{,}6}{100}\right] = 0{,}99$$

$P(0{,}274 < p \leq 0{,}326) = 0{,}99$

Il y a 99 % de chances que p soit comprise entre 27,4 % et 32,6 %.

EXERCICE N° 125

X

Un fabricant reçoit de son fournisseur habituel une livraison de pièces dont il veut contrôler la longueur. La dimension X suit une loi inconnue de moyenne m et d'écart-type σ inconnus. Il extrait un échantillon de 6 pièces qui présentent les dimensions suivantes :

50 40 45 43 47 45

Estimer à partir de cet échantillon la variance σ^2 de la longueur des pièces reçues, et déterminer, avec un risque de 10 %, un intervalle de confiance pour la longueur moyenne m.

On donne $\sum x_i = 270$ et $\sum x_i^2 = 12\,208$

(Licence Sciences économiques 2ᵉ année - Octobre 1977)

Solution

Estimation de la variance σ^2.

Un bon estimateur de σ^2 est s^2 car la taille de l'échantillon est petite

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

$$\bar{x} = \frac{270}{6} = 45$$

$$s^2 = \frac{12\,208 - 6 \times (45)^2}{5} = \frac{12\,208 - 12\,150}{5} = \frac{58}{5} = 11{,}6$$

La variance serait égale à 11,6 et l'écart-type à 3,40.

Encadrement de m à 90 %

n étant très petit, la variable centrée réduite

$$t = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

suit une loi de Student-Fisher à 5 degrés de liberté

$$\left.\begin{aligned} \alpha = 0{,}90 \Rightarrow 1 - \alpha = 0{,}10 \\ V = n - 1 = 5 \end{aligned}\right\} \Rightarrow t = 2{,}015$$

$$P\left[45 - 2{,}015 \times \frac{3{,}4}{2{,}4} < m \leq 45 + 2{,}015 \times \frac{3{,}4}{2{,}4}\right] = 0{,}90$$

$$P\,(45 - 2{,}85 < m \leq 45 + 2{,}85) = 0{,}90$$

$$P\,(42{,}15 < m \leq 47{,}85) = 0{,}90$$

Il y a 90 chances sur 100 que la longueur moyenne des pièces reçues soit comprise entre 42,15 et 47,85.

EXERCICE N° 126

La durée de vie d'un certain type de moteur de voiture, mesurée par le nombre de milliers de kilomètres parcourus, est une variable aléatoire X qui suit une loi normale de moyenne m et d'écart-type σ inconnus.

Un échantillon de 125 moteurs a donné les résultats suivants :

$$\bar{x} = \frac{\sum x_i}{125} = 70$$

$$\sum (x_i - \bar{x})^2 = 28\,125$$

1°) Estimer les paramètres m et σ de la loi de X.

2°) La loi de X étant exactement celle déterminée précédemment, quelle est la loi de la moyenne de l'échantillon précédent ?

(Licence Sciences économiques, 2ᵉ année - Octobre 1978)

Solution

1°) *Estimation de la moyenne et l'écart-type de la population.*

Un bon estimateur de la moyenne m de la population est $\bar{x}$ de l'échantillon.

Or

$$\bar{x} = \frac{\sum x_i}{125} = 70$$

La durée de vie moyenne du moteur serait alors de 70 milles kilomètres.

La variance σ^2 est valablement estimée par s^2

$$\sigma^2 \simeq s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{28\,125}{124} = 226{,}81$$

$$\sigma \simeq 15{,}06$$

Remarque : La taille de l'échantillon étant assez grande, la variance de l'échantillon peut également donner une bonne estimation de celle de la population

$$\sigma^2 \simeq \sigma_e^2 = \frac{28\,125}{125} = 225$$

$$\sigma = 15$$

2°) *Loi de* $\bar{x}$:

Le caractère X étant gaussien, $\bar{x}$ suit également une loi normale (théorème central-limit)

Ses paramètres sont :

$$E(\bar{x}) = m = 70$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{125}} = \frac{15}{11{,}18} = 1{,}34$$

$$\bar{x} \rightarrow N(70\,;\,1{,}34)$$

EXERCICE N° 127

On admet que la durée de vie X d'un pneumatique est une variable aléatoire normale

$$N(m\,;\,\sigma)$$

m et σ étant inconnus.

On choisit au hasard 25 pneus que l'on fait rouler jusqu'à usure complète. On désigne par $X_1, \ldots, X_2, \ldots x_n$ les nombres de kilomètres parcourus. Les X_i sont supposés indépendantes.

On pose

$$\bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i$$

$$\text{et } s^2 = \frac{1}{24}\sum_{i=1}^{25} (x_i - \bar{x})^2$$

On suppose que :

$$\bar{x} = 2 \cdot 10^4 \text{ km} \quad \text{et} \quad s^2 = 25 \cdot 10^6$$

1°) Estimer m par un intervalle de confiance à 0,95. On justifiera au préalable, par un petit raisonnement, la forme de l'intervalle de confiance en partant d'une quantité aléatoire dont on précisera la loi.

2°) A partir de l'échantillon considéré, donner une limite à la valeur empirique de la variance σ^2 avec une probabilité égale à 0,80.

3°) On désigne par p la probabilité pour qu'un pneu roule plus de 15 000 km, p étant inconnue. On désigne par f la fréquence relative expérimentale des pneus ayant roulé plus de 15 000 km dans l'échantillon de 25 pneus.

On suppose que f = 0,60

a) En partant d'une quantité aléatoire dont on précisera la nature, établir la loi de probabilité de f. Donner ses paramètres.

b) En déduire un intervalle de confiance de p. Calculer par une méthode appropriée cet intervalle pour un degré de confiance égal à 0,95.

c) Dans le cas où la taille de l'échantillon serait n, démontrer que f est un bon estimateur de p et en déduire le nouvel intervalle de confiance à 0,95.

d) Si on veut connaître l'estimation de p au $\frac{1}{100}$ près avec une probabilité de 0,99, quelle doit être la taille de l'échantillon ?

(Licence Sciences économiques 2e année - Juin 1978)

Solution

1°) *Intervalle de confiance à 95 % pour m :*

Le caractère X, durée de vie d'un pneumatique, étant gaussien on en déduit que :

$$\bar{x} \to N\left(m ; \frac{\sigma}{\sqrt{n}}\right)$$

Il en résulte que :

$$P\left[m - t\frac{\sigma}{\sqrt{n}} < \bar{x} \leq m + t\frac{\sigma}{\sqrt{n}}\right] = \int_{-t}^{+t} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$= 2\Pi(t) - 1$$

Par ailleurs

$$m - t\frac{\sigma}{\sqrt{n}} < \bar{x} \Rightarrow m > \bar{x} + t\frac{\sigma}{\sqrt{n}}$$

$$m + t\frac{\sigma}{\sqrt{n}} > \bar{x} \Rightarrow m \geq \bar{x} - t\frac{\sigma}{\sqrt{n}}$$

Par conséquent les deux événements

$$\left[m - t\frac{\sigma}{\sqrt{n}} < \bar{x} \leq m + t\frac{\sigma}{\sqrt{n}}\right] \text{ et } \left[\bar{x} - t\frac{\sigma}{\sqrt{n}} < m \leq \bar{x} + t\frac{\sigma}{\sqrt{n}}\right]$$

sont identiques et ont ainsi la même probabilité, d'où l'intervalle de confiance à 95 %

$$P\left[\bar{x} - t\frac{\sigma}{\sqrt{n}} < m \leq \bar{x} + t\frac{\sigma}{\sqrt{n}}\right] = 2\Pi(t) - 1 = \alpha$$

$\alpha = 0,95 \Rightarrow t = 1,96$

$\bar{x} = 20\,000$

σ peut être estimé par s :

$s = \sqrt{25 \cdot 10^6} = 5 \cdot 10^3 = 5\,000$

$$P\left[20\,000 - 1,96\frac{5\,000}{\sqrt{25}} < m \leq 20\,000 + 1,96\frac{5\,000}{\sqrt{25}}\right] = 0,95$$

$P\,(20\,000 - 1\,960 < m \leq 20\,000 + 1\,960) = 0,95$

$P(18\,040 < m \leq 21\,960) = 0,95$

2°) *Limite de la valeur empirique de la variance σ^2*

La quantité khi-deux (χ^2) est donnée par l'expression suivante :

$$\chi^2 = n \cdot \frac{\sigma_e^2}{\sigma^2}$$

$$\left.\begin{array}{l} P\,(\chi^2 > a) = 0,80 \\ \nu = n - 1 = 24 \end{array}\right\} \Rightarrow a = 18,062$$

d'où :

$$n\frac{\sigma_e^2}{\sigma^2} > 18,062$$

$$n\frac{\sigma_e^2}{18,062} > \sigma^2$$

$$\sigma^2 < n\frac{\frac{\sum (x_i - \bar{x})^2}{n}}{18,062}$$

$$\sigma^2 < \frac{\sum (x_i - \bar{x})^2}{18,062}$$

or :

$$\frac{1}{24}\sum (x_i - \bar{x})^2 = 25 \cdot 10^6$$

$$\sum (x_i - \bar{x})^2 = 600 \cdot 10^6$$

$$\sigma^2 < \frac{600 \cdot 10^6}{18,062}$$

$$\sigma^2 < 33,2189 \cdot 10^6$$

$$\sigma^2 < 332 \cdot 10^5$$

3°) *Etude de f et p*

a) *Loi de f*

La durée de vie X du pneumatique peut être soit supérieure soit inférieure à 15 000 km. Elle suit une loi binomiale de paramètres

$n = 25$ et p inconnue.

Cette loi peut être approchée par une loi normale de paramètres

np et $\sqrt{npq}$

d'où :

$$t = \frac{X - np}{\sqrt{npq}} \to N(0,1)$$

$$t = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} \to N(0,1)$$

$$t = \frac{f - E(f)}{\sigma f} \to N(0,1)$$

Par conséquent :

$$f \to N\left(p\,;\sqrt{\frac{pq}{n}}\right)$$

b) *Intervalle de confiance pour p :*

Par un raisonnement analogue à celui qui a permis d'établir l'intervalle de confiance de m, on déduit que :

$$P\left[f - t\sqrt{\frac{pq}{n}} < p \leq f + t\sqrt{\frac{pq}{n}}\right] = \int_{-t}^{+t} \frac{1}{\sqrt{2\Pi}}\, e^{-\frac{u^2}{2}} du = 2\Pi(t) - 1$$

puisque p est une inconnue, on peut estimer pq par le maximum possible soit $\frac{1}{4}$; d'où

$$P\left[f - t\sqrt{\frac{1}{4n}} < p \leq f + t\sqrt{\frac{1}{4n}}\right] = 2\Pi(t) - 1 = \alpha$$

$\alpha = 0{,}95 \Rightarrow t = 1{,}96 \# 2$

$$P\left[0{,}60 - 2\sqrt{\frac{1}{4 \times 25}} < p \leq 0{,}60 + 2\sqrt{\frac{1}{4 \times 25}}\right] = 0{,}95$$

$P(0{,}40 < p \leq 0{,}80) = 0{,}95$

c) *Estimation de p par f :*

On sait que :

$E(f) = p$

$\sigma f = \frac{pq}{n} \to 0$ si $n \to \infty$

f est donc un estimateur correct de p ; d'où

$$P\left[f - t\sqrt{\frac{f(1-f)}{n}} < p \leq f + t\sqrt{\frac{f(1-f)}{n}}\right] = 0{,}95$$

$P(0{,}60 - 2 \times 0{,}098 < p \leq 0{,}60 + 2 \times 0{,}098) = 0{,}95$

$P(0{,}600 - 0{,}196 < p \leq 0{,}600 + 0{,}196) = 0{,}95$

$P(0{,}404 < p \leq 0{,}796) = 0{,}95$

d) *Taille de l'échantillon*

Nous savons (cf. exercices précédents) que

$$t\sqrt{\frac{pq}{n}} \leq \varepsilon \Leftrightarrow t^2\,\frac{pq}{n} \leq \varepsilon^2 \Leftrightarrow n \geq \frac{t^2 p(1-p)}{\varepsilon^2}$$

$$n \geq \frac{(2{,}58)^2 \times (0{,}60) \times (0{,}40)}{10^{-4}} \Rightarrow n \geq 15\,975$$

on retiendra la valeur 16 000

EXERCICE N° 128

Une entreprise produit des fils de métal. En moyenne la charge de rupture d'une certaine catégorie de fils est 851 g avec un écart-type de 75 g.

A. On prélève un échantillon de 100 fils.

1. que peut-on dire de la loi de la moyenne des charges de rupture dans cet échantillon ?

2. en déduire les intervalles d'encadrement de cette moyenne à 5 % et à 1 % de risque.

B. Les cent expériences de rupture en charge d'un fil sont faites sur un échantillon de fils prélevés sur l'ensemble de la production d'une entreprise.

1. les 100 expériences ont donné les résultats suivants :

$$\sum_{1}^{100} x_i = 84\,707 \text{ g} \qquad \Bigg| \qquad \sum_{1}^{100} x_i^2 = 72\,413\,389$$

x_i étant la charge de rupture du fil i, i allant de 1 à 100.

Quel poids maximum peut-on affirmer que ces fils peuvent supporter avec un risque d'erreur de 1 % ? Etablir l'intervalle de confiance à 99 % de l'estimation de la charge moyenne de rupture m.

2. Supposons qu'au lieu de prélever 100 fils pour expérimenter la charge de rupture, on en ait seulement choisi 12, qui ont donné une charge moyenne de 854,17 g avec un écart-type de 67,61 g.

Donner l'intervalle de confiance à 99 % de la charge moyenne de rupture m.

(Licence ès-sc. Eco. 2e année - Octobre 1970)

Solution

A. Expérience sur un échantillon de 100 fils :

1. Loi de $\bar{x}$:

$\bar{x}$ est la charge moyenne de rupture.

La taille de l'échantillon étant assez grande ($n = 100$), on peut affirmer que $\bar{x}$ suit une loi de Laplace-Gauss.

$$\bar{x} \to N\left(m , \frac{\sigma}{\sqrt{n}}\right)$$

m et σ étant la moyenne et l'écart-type de la population totale

$E(\bar{x}) = m = 851$

$\sigma_{\bar{x}} = \frac{75}{10} = 7,5$

$\bar{x} \to N(851 ; 7,5)$.

2. Intervalles d'encadrement de $\bar{x}$:

a) à 5 % de risque :

$1 - \alpha = 0,95 \Rightarrow 2\Pi(t) - 1 = 0,95 \Rightarrow \Pi(t) = 0,975$

$\Rightarrow t = 1,96$ (table de la loi normale).

$P(851 - 1,96 \times 7,5 < \bar{x} \leq 851 + 1,96 \times 7,5) = 0,95$

$P(836,3 < \bar{x} \leq 865,7) = 0,95$

b) à 1 % de risque :

$1 - \alpha = 0,99 \Rightarrow 2\Pi(t) - 1 = 0,99 \Rightarrow \Pi(t) = 0,995 \Rightarrow t = 2,58$

$P(851 - 2,58 \times 7,5 < \bar{x} \leq 851 + 2,58 \times 7,5) = 0,99$

$P(831,65 < \bar{x} \leq 870,35) = 0,99$.

B. Etude de m, moyenne de la population :

1. Calcul de la charge maximum et des intervalles de confiance de m :

Il s'agit de trouver une valeur m, charge moyenne que ces fils peuvent supporter avec une probabilité de 99 %. Pour ce faire il faut calculer au préalable $\bar{x}$, estimer l'écart-type de la population (σ) et en déduire celui de la moyenne ($\sigma_{\bar{x}}$)

$$\bar{x} = \frac{84\,797}{100} = 847,97$$

$$\sigma_e^2 = \frac{72\,413\,389}{100} - (847,97)^2 = 5\,080,7$$

$$\sigma^2 = s^2 = \frac{n}{n-1}\,\sigma_e^2 = \frac{5\,080,7 \times 100}{99}$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{5\,080,7 \times 100}{99 \times 100}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{5\,080,7}{99}} = \sqrt{51,32} = 7,16$$

Il convient maintenant de déterminer m tel que

$P(m + t\,\sigma_{\bar{x}} < \bar{x}) = 0,99$

$$\left.\begin{array}{l} P\left(t < \frac{\bar{x} - m}{\sigma_{\bar{x}}}\right) = 0,99 \\ \Pi(2,33) = 0,99 \end{array}\right\} \Rightarrow \frac{\bar{x} - m}{\sigma_{\bar{x}}} = 2,33$$

$\bar{x} - m = 2,33\,\sigma_{\bar{x}}$

$m = \bar{x} - 2,33\,\sigma_{\bar{x}} = 847,97 - 2,33 \times 7,16$

$m = 831,29$

Il y a donc une chance sur 100 que la moyenne de la population soit inférieure à 831,29 g.

Intervalle de confiance à 99 % :

$P(847,97 - 2,58 \times 7,16 < m \leq 847,97 + 2,58 \times 7,16) = 0,99$

$P(847,97 - 18,47 < m \leq 847,97 + 18,47) = 0,99$

$P(829,5 < m \leq 866,44) = 0,99$

Il y a 99 chances sur 100 que la charge moyenne dans la population soit comprise entre 830 et 866.

2°) Cas d'un échantillon de taille n = 12

$$P\left(\bar{x} - t\frac{s}{\sqrt{n}} < m < \bar{x} + t\frac{s}{\sqrt{n}}\right) = 0,99$$

$$\left.\begin{array}{ll} \alpha = 0,99 & \Rightarrow 1 - \alpha = 0,01 \\ n = 12 & \Rightarrow \nu = n - 1 = 11 \end{array}\right\} \Rightarrow |t| > 3,106$$

(lecture de la table de Student-Fisher)

$\bar{x} = 857,17$ g.

$$s = \sigma_e\sqrt{\frac{n}{n-1}} = 67,61\sqrt{\frac{12}{11}} = 70,62$$

$$P\left(857,17 - 3,106 \times \frac{70,62}{\sqrt{12}} < m \leq 857,17 + 3,106 \cdot \frac{70,62}{\sqrt{12}}\right) = 0,99$$

$P(857,17 - 3,106 \times 20,41 < m \leq 857,17 + 3,106 \times 20,41) = 0,99$

$P(790,78 < m \leq 917,56) = 0,99$

$791 < m \leq 917$

L'intervalle trouvé ici est plus large que celui trouvé auparavant : en effet on gagne en précision en agrandissant l'échantillon.

Solution

1°) Intervalle de confiance à 0,99 dans le cas d'un échantillon de taille $n = 100$:

$$\left.\begin{array}{l} n = 100 \\ k = 5 \end{array}\right\} \Rightarrow f = \frac{5}{100} = 0{,}05$$

$$P\left(f - t\sqrt{\frac{f(1-f)}{100}} < p \leq f + t\sqrt{\frac{f(1-f)}{100}}\right) = 0{,}99$$

Pour un seuil de confiance de 0,99 (risque de 1 %) la lecture de la table normale donne $t = 2{,}58$ (on sait que $2\pi(t) - 1 = 0{,}99$).

$$P\left(0{,}05 - 2{,}58\sqrt{\frac{0{,}05 \times 0{,}95}{100}} < p \leq 0{,}05 + 2{,}58\sqrt{\frac{0{,}05 \times 0{,}95}{100}}\right) = 0{,}99$$

$$P\left(0{,}05 - 2{,}58\sqrt{4{,}75 \cdot 10^{-4}} < p \leq 0{,}05 + 2{,}58\sqrt{4{,}75 \cdot 10^{-4}}\right) = 0{,}99$$

$$P\left(0{,}05 - 2{,}58 \times 2{,}17 \cdot 10^{-2} < p \leq 0{,}05 + 2{,}58 \times 2{,}17 \cdot 10^{-2}\right) = 0{,}99$$

$$P(0{,}050 - 0{,}056 < p \leq 0{,}050 + 0{,}056) = 0{,}99$$

$$P(0{,}01 < p \leq 0{,}14) = 0{,}99$$

Ainsi malgré la taille 100 de l'échantillon, celui-ci ne donne pas une bonne précision pour l'intervalle de confiance de p en raison notamment du nombre réduit des pièces défectueuses observées.

2°) Détermination de la taille n en fonction de la précision souhaitée :

$$P\left(f - t\sqrt{\frac{f(1-f)}{n}} < p \leq f + t\sqrt{\frac{f(1-f)}{n}}\right) = 0{,}99$$

$$P\left(-t\sqrt{\frac{f(1-f)}{n}} < p - f \leq t\sqrt{\frac{f(1-f)}{n}}\right) = 0{,}99$$

$$\alpha = 0{,}99 \Rightarrow t = 2{,}58$$

$$P\left(-2{,}58\sqrt{\frac{f(1-f)}{n}} < p - f \leq 2{,}58\sqrt{\frac{f(1-f)}{n}}\right) = 0{,}99$$

ou encore :

$$P\left(|p - f| < 2{,}58\sqrt{\frac{f(1-f)}{n}}\right) = 0{,}99$$

a) $|p - f| = |i| = 0{,}5\ \% = \frac{0{,}5}{100} = 0{,}005 = 5 \cdot 10^{-3}$

f demeure égale à 0,05. D'où :

$$2{,}58\sqrt{\frac{0{,}05 \times 0{,}95}{n}} = 5 \cdot 10^{-3}$$

$$\frac{(2{,}58)^2 \times (0{,}05) \times (0{,}95)}{n} = 25 \cdot 10^{-6}$$

$$n = \frac{(2{,}58)^2 \times (5 \cdot 10^{-2})(95 \cdot 10^{-2})}{25 \cdot 10^{-6}} = 12\,647$$

L'intervalle de confiance qui en résulte est, après calcul : $(0{,}945 < p \leq 0{,}955)$. Ce qui fournit une meilleure précision.

b) Pour un intervalle de confiance 2 fois plus petit il faut un échantillon d'une taille 4 fois plus grande, soit : $12\,647 \times 4 = 50\,588$.

c) La taille de la population n'intervient pas dans la détermination de la taille de l'échantillon pour rechercher une meilleure précision de l'estimation.